Understanding global model systematic shortwave radiation errors in subtropical marine boundary layer cloud regimes

Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2018MS001346

Understanding Global Model Systematic Shortwave Radiation Errors in Subtropical Marine Boundary Layer Cloud Regimes

Maike Ahlgrimm[1], **Richard M. Forbes**[1], **Robin J. Hogan**[1], **and Irina Sandu**[1]

[1] European Centre For Medium-Range Weather Forecasts, Shinfield Park, UK

Key Points:
- A methodology to attribute the shortwave radiation bias to specific errors in cloud properties is applied to the global ECMWF model
- In trade cumulus cloud, the shortwave bias is almost entirely due to errors in the model's liquid water path
- In stratocumulus clouds, errors in cloud cover, water path, effective radius, and subgrid heterogeneity all contribute to the shortwave bias

Correspondence to:
M. Ahlgrimm,
maike.ahlgrimm@ecmwf.int

Citation:
Ahlgrimm, M., Forbes, R., Hogan, R. J., & Sandu, I. (2018). Understanding global model systematic shortwave radiation errors in subtropical marine boundary layer cloud regimes. *Journal of Advances in Modeling Earth Systems*, *10*, 2042–2060. https://doi.org/10.1029/2018MS001346

Received 19 APR 2018
Accepted 29 JUN 2018
Accepted article online 10 JUL 2018
Published online 24 AUG 2018

Abstract Global numerical weather prediction and climate models are subject to long-standing systematic shortwave radiation errors due to deficiencies in the representation of boundary layer clouds over the ocean. In the subtropics, clouds are typically too reflective in the cumulus regime and not reflective enough in the stratocumulus regime. Potential sources of error include cloud cover, liquid water path, effective radius, and subgrid heterogeneity, but diagnosing the absolute contributions of each to the radiation bias is hampered by uncertainties and sometimes contradictory information from different observational products. This paper draws on a set of ship-based observations of boundary layer clouds obtained during the ARM MAGIC campaign along a northeast Pacific Ocean transect, crossing both stratocumulus and shallow cumulus cloud regimes. The surface-based observations of cloud properties are compared with various satellite products, taking account of the diurnal cycle, to provide an improved quantitative assessment of the deficiencies in the European Centre for Medium-Range Weather Forecasts global numerical weather prediction model. A series of off-line radiation calculations are then performed to assess the impact on the shortwave radiation bias of correcting each of the model's deficiencies in cloud characteristics along the transect. A reduction in the bias is achieved by improving the agreement between modeled and observed in-cloud liquid water path frequency distributions. In the cumulus regime, this is accomplished primarily by reducing the all-sky water path, while for the stratocumulus regime, an underestimate of cloud cover and liquid water and an overestimate in effective radius and subgrid heterogeneity all contribute to a lack of reflected shortwave radiation.

1. Introduction

The radiative impact of clouds is of fundamental importance for weather forecasting and climate projections, yet there are certain regime-dependent systematic radiation errors that are common and long standing across many different global modeling systems (Li et al., 2013). In particular, boundary layer cloud is prevalent over the oceans and due to the contrast in albedo between clouds and the underlying ocean surface, model biases in the representation of these clouds have a strong impact on the shortwave (SW) radiation. Indeed, marine boundary layer (MBL) clouds remain one of the major sources of uncertainty in climate projections (Bony & Dufresne, 2005; Vial et al., 2013; Webb et al., 2013). Most commonly, SW biases with opposite sign are found in stratocumulus and trade cumulus regimes over the subtropical oceans (Kalmus et al., 2014; Li et al., 2013; Teixeira et al., 2011). For the stratocumulus regions, reflected SW radiation is underestimated, typically due to an underprediction of cloud cover (Lin et al., 2014), although there may be other contributing factors. In the trade cumulus regions, where too much SW is reflected, the cause for the bias is not always clear. Nam et al. (2012) show that many of the CMIP5 climate models overestimate the albedo in the cumulus regime while also apparently underestimating cloud occurrence—the *too few*, *too bright* problem. This indicates that there must be compensating errors in cloud occurrence, condensate amount, and microphysical properties of the cloud to produce the resulting overestimate of reflected SW radiation.

Using a variety of observations to simultaneously constrain the different aspects of the model predicted cloud field is a necessary step in the process of improving the model parameterizations, but this is not always so straightforward due to uncertainties and ambiguities in the observations. For example, attributing the cause of the radiation bias in the trade cumulus regime is difficult because of the large uncertainties still associated with satellite-based retrievals of liquid water path (LWP). Li et al. (2008) show that discrepancies between LWP estimates are large between different satellite observational products as well as between various global models and analyses. Seethala and Horváth (2010) find that LWP retrievals from passive

microwave instruments using the Wentz algorithm (Hilburn & Wentz, 2008; Wentz & Spencer, 1998) are subject to retrieval biases particularly in areas with broken cloud cover, such as those found in the trade cumulus regions. Studies comparing satellite retrievals based on microwave radiometer data with retrievals based on optical depth/effective radius (Clouds and the Earth's Radiant Energy System/Moderate-resolution Imaging Spectroradiometer, CERES/MODIS) conclude that agreement is good where cloud cover is overcast and nonprecipitating (Bennartz, 2007; Greenwald, 2009; Painemal et al., 2016; Seethala & Horváth, 2010) but diverge for broken cloud cover and in the presence of precipitation (Greenwald et al., 2018; Lebsock & Su, 2014).

Similar to other global models, the Integrated Forecast System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF) also underestimates the SW albedo in the marine stratocumulus regime and overestimates it in the trade cumulus regime. This pattern is present in long integrations at low resolution (Figure 1a, a small ensemble of 4-year-long, free-running, uncoupled experiments with a grid resolution of 80 km) as well as close to the analysis in high-resolution forecasts (Figure 1b, first 24 hr of operational forecasts with a grid resolution of 9 km). The SW radiation bias in the subtropics is very similar across forecast time scales and model resolutions.

At first glance, when considering annual or seasonal means for cloud cover averaged across the entire day, the too bright bias seen in the IFS for the trade cumulus regime also appears to coincide with too few clouds as seen in Nam et al. (2012). In this study, we take advantage of ship-based observations gathered in the northeast Pacific and a variety of satellite products to better understand the sources of these SW radiation biases in MBL clouds. We aim to attribute the causes of the radiation bias by quantifying the contributions of different factors including cloud fraction, cloud water path, effective radius, vertical overlap, cloud heterogeneity, and the impact of 3-D radiative effects. The fact that these radiation biases are already present after a few hours in the model forecasts facilitates this task. The model is still close to the observed state in short-range forecasts, making it easier to attribute model errors to individual factors compared to longer time ranges where process interactions can obscure the causes of the bias. We estimate the contribution of the various sources of model error through a combination of short-range forecasts, surface-based and satellite observations, and an off-line version of the IFS radiation scheme (ECRAD; Hogan & Bozzo, 2016).

Section 2 describes the observations and model used in this study. The different observational products and the model are then compared in section 3. Section 4 describes the off-line radiation experiments, and conclusions are drawn in section 5.

2. Observations and IFS Model

2.1. Satellite Observations

The CERES Energy Balanced and Filled (EBAF, Edition 4; Loeb et al., 2018) data products are used to provide top-of-atmosphere (TOA) SW radiative fluxes. Climatological TOA-fluxes from this product are averaged across all times of day, as well as specified periods (months, years). The MODIS product used in this study is the combined CERES-MODIS SSF 1° retrieval (Ed4A; Loeb et al., 2018), which provides a daytime liquid cloud cover (used here as LCC equivalent) alongside a LWP retrieval based on optical depth and effective radius (VNIR and 3.7-μm channels), assuming adiabatic cloud water concentration within the cloud layer. The CERES-MODIS SSF product uses MODIS-calibrated radiances (available from Goddard Space Flight Center as part of Collection 5), while cloud properties are retrieved using the algorithms described in Minnis, Sun-Mack, Chen, et al. (2011) and Minnis, Sun-Mack, Young, et al. (2011). The Multi-Sensor Advanced Climatology (MAC) of LWP (Elsaesser et al., 2017) also provides an all-day mean LWP estimate by combining retrievals from the available satellite microwave sensors in any given period to provide a best estimate climatology covering a 29-year period. The MAC-LWP product (Elsaesser et al., 2016) uses an updated version of the Wentz algorithm (Wentz, 2013) to retrieve LWP from microwave radiances. An advantage of combining several microwave sensors in the MAC-LWP product is that the diurnal cycle is better resolved than in the CERES-MODIS SSF products, which can provide retrievals only at specific overpass times.

A second cloud cover product independent from CERES-MODIS SSF is the GCM Oriented CALIPSO Cloud Product (GOCCP, v3.1.2; Chepfer et al., 2010; Guzman et al., 2017) derived from CALIPSO backscatter. CALIPSO (Winker et al., 2003) is part of the A-train constellation and thus observes clouds in the early afternoon and late night hours (around 1:30 a.m. and 1:30 p.m. solar time, ST). Table 1 lists the details of how, when, and for what time range data products were accessed.

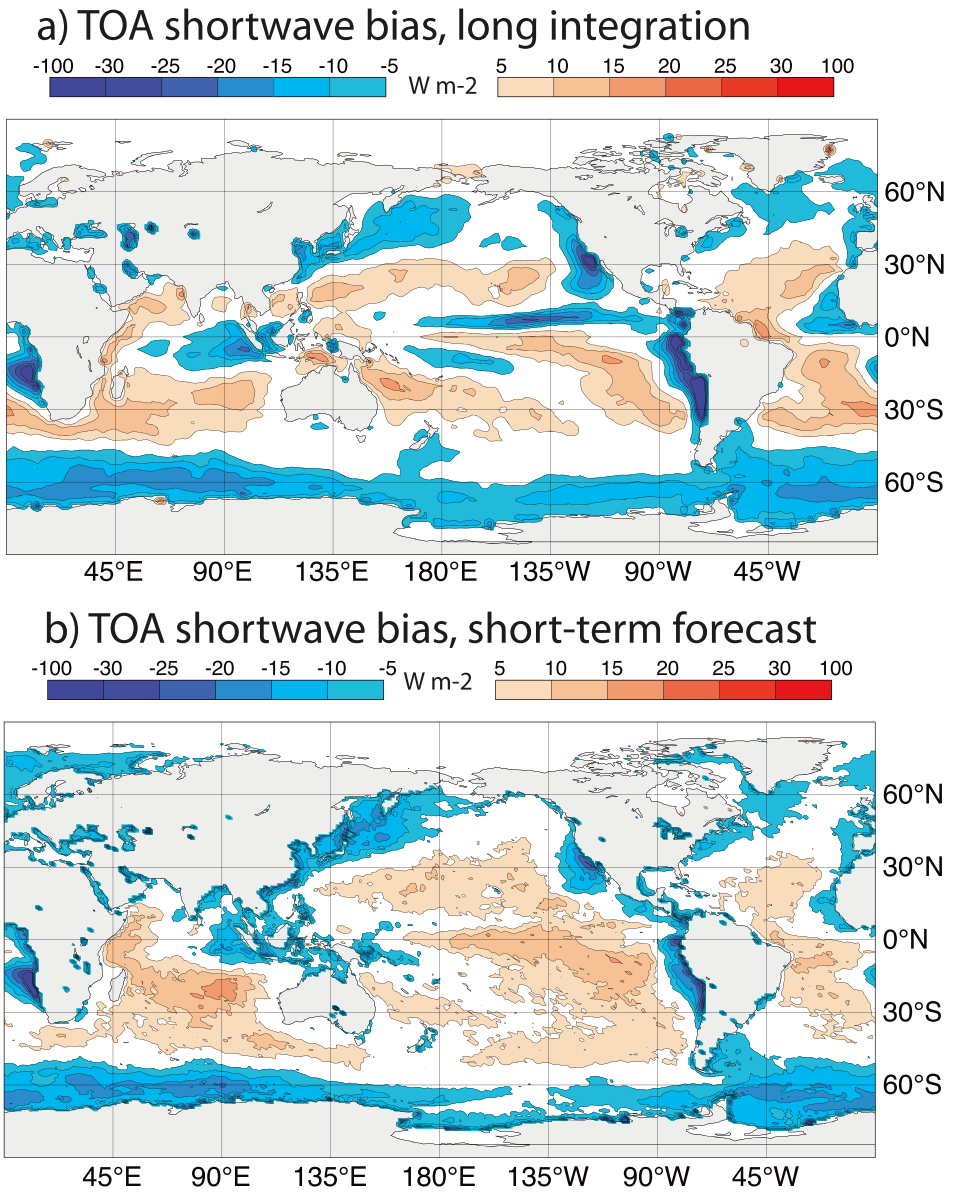

Figure 1. Top-of-atmosphere upwelling shortwave radiation bias from the Integrated Forecast System (CY43R1) compared to Clouds and the Earth's Radiant Energy System Energy Balanced and Filled observations for (a) year-long free-running ensemble run averaged over years 2000–2004 (model climate) and (b) operational short-range (24 hr) forecasts averaged for the year October 2016 to September 2017. Blue areas indicate regions where too little radiation is reflected back to space, while red areas indicate regions with an overestimated shortwave albedo. (Sign convention used here is positive for upwelling fluxes.) Shortwave fluxes are averaged across all times of day. TOA = top of atmosphere.

2.2. The Ship-Based MAGIC Observational Campaign

The Marine ARM GPCI Investigation of Clouds (MAGIC) was an observational campaign during the summer of 2013 where a range of active and passive instruments were placed on a container ship traveling regularly between Los Angeles, California, and Hawaii (Lewis et al., 2012; Zhou et al., 2015). The MAGIC transect crosses an area of both stratocumulus and cumulus boundary layer clouds, which coincides with a persistent radiation bias in the model and is therefore an appropriate data set for this investigation. Through the months of June, July, and August (JJA) of 2013, the ship performed eight round trips during which good quality observations of cloud and aerosol properties were gathered (legs 10–17). We use the ceilometer-detected cloud base to provide a LCC estimate, and LWP retrieved from the three-channel microwave radiometer (Cadeddu et al., 2013; Fielding et al., 2015). Given the large uncertainty in water path estimates from satellite-borne instruments, the

Table 1
Data Products

Product name	DOI or weblink	Date range used	Accessed on
CERES-EBAF Ed4	10.5067/TERRA+AQUA/CERES/EBAF-TOA_L3B004.0	2000–2004, JJA 2013, 2016/10 to 2017/9	2018/2/9
CERES-MODIS SSF 1deg Ed4A	10.5067/AQUA/CERES/SSF1DEGMONTH_L3.004A	JJA 2006–2016	2017/5/30
GOCCP MapLowMid-High_2×2×L40 v3.1.2	http://climserv.ipsl.polytechnique.fr/cfmip-obs/	JJA 2006–2016	2018/2/14
GOCCP 3D_CloudFraction_OPAQ _2×2×L40 v3.1.2	http://climserv.ipsl.polytechnique.fr/cfmip-obs/	JJA 2006-2016	2018/5/28
MAC-LWP	10.5067/MEASURES/MACLWPM	JJA 2013	2017/11/6
MAGIC, 3ch MWR retrieval	https://doi.org/10.5439/1435798	JJA 2013	2016/10/12
MAGIC, ceilometer	https://www.archive.arm.gov/discovery/\#v/datastreams/s/streams::magceilM1.b1	JJA 2013	2017/2/10
ECMWF model data along MAGIC track	https://iop.archive.arm.gov/arm-iop/2012/ mag/magic/wagener-ecmwf/	JJA 2013	2014/12/1

Note. Dates are formatted as year/month/day. CERES = Clouds and the Earth's Radiant Energy System; EBAF = Energy Balanced and Filled; MODIS = Moderate-resolution Imaging Spectroradiometer; MAGIC = Marine ARM GPCI Investigation of Clouds; ECMWF = European Centre for Medium-Range Weather Forecasts; LWP = liquid water path; JJA = June-July-August.

surface-based observations from MAGIC offer an independent set of observations with a surface perspective and higher temporal and spatial resolution than from satellite. Details for data access can be found in Table 1.

The ship also carried instruments to measure near-surface aerosol properties, including an Ultra-High Sensitivity Aerosol Spectrometer (UHSAS). The emissions from the ship's exhaust stack posed a particular challenge for aerosol measurements. McGibbon and Bretherton (2017) found that a reasonable cloud droplet number concentration (CDNC) could be derived from the UHSAS by counting all particles exceeding 98.8 nm as cloud condensation nuclei (CCN) disregarding any periods with high concentrations of small particles to avoid stack contamination and converting this near-surface concentration to cloud-level CDNC with a simple parametric relationship to match coincident CDNC retrieved from GOES imagery. Here we use the hourly time series of CCN produced by and described in McGibbon and Bretherton (2017) and use the same parametric relationship described in their study to derive CDNC along the MAGIC transect.

2.3. IFS Global Model

The IFS is a global model producing operational medium- and extended-range forecasts at ECMWF. The version used here is Cycle 43r1 that was operational from 22 November 2016 to 10 July 2017. As shown in Figure 1, the same bias patterns found in year-long IFS integrations are already present within the first day of the forecast. Analyzing short-range forecasts has the advantage that the model state is still close to the observed state, allowing a direct comparison with the MAGIC observations and making it somewhat easier to identify forcing and response before slower feedbacks and process interactions have had time to act. Hindcasts with the IFS were run at a horizontal grid resolution of approximately 25 km (cubic octahedral reduced Gaussian grid with truncation at wave number 399) and with 137 vertical levels, initialized daily at 12 UTC for the JJA 2013 season. For comparison with satellite observations, 3-hourly output from 12- to 33-hr forecast lead times is averaged for the season and interpolated onto a 1° latitude-longitude grid to match the typical data product resolution. All model data are matched to the observation time and locations as close as possible given the time and space discretization of the output. For direct comparison with MAGIC observations, hourly model data from daily forecasts with a lead time of 12–35 hr are stitched together to provide a continuous hourly time series for the JJA 2013 season, which is then matched in time and space with the ship's location using the nearest grid points to the ship track.

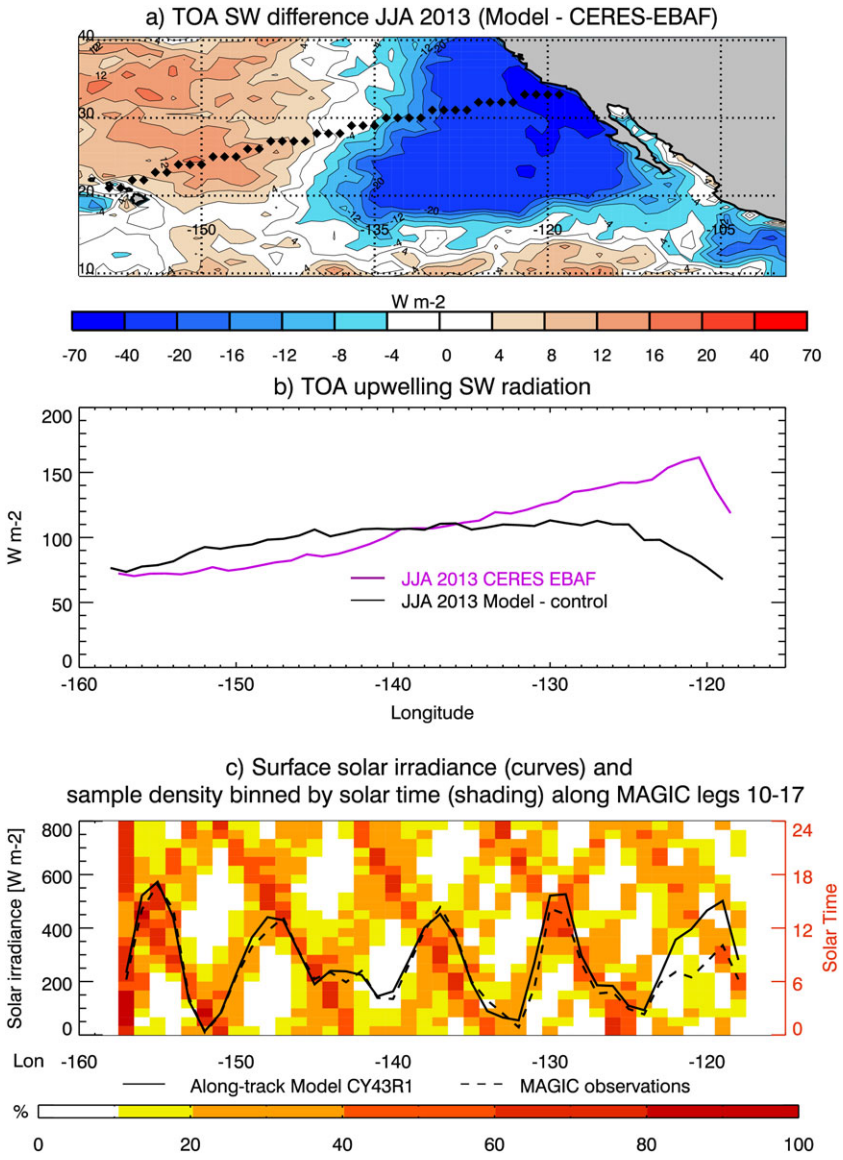

Figure 2. (a) Map of TOA SW bias versus CERES EBAF for JJA 2013 over the northeast Pacific. Diamonds mark the nearest model grid point along the average MAGIC track. (b) Model and CERES-EBAF TOA upwelling SW averaged for JJA 2013 along MAGIC track. Radiative fluxes in panels (a) and (b) are averaged over all times of day. (c) Surface solar irradiance composited across MAGIC Legs 10–17 from model and MAGIC observations matched in time and space (black curves). Density of solar irradiance samples (normalized by maximum number of samples in any bin) for each longitude and solar time bin (shading). TOA = top of atmosphere; SW = shortwave; CERES = Clouds and the Earth's Radiant Energy System; EBAF = Energy Balanced and Filled; MAGIC = Marine ARM GPCI Investigation of Clouds; ARM = Atmospheric Radiation Measurement; GPCI = GCSS Pacific Cross-section Intercomparison.

3. Comparison With Observations

Figure 2a shows that the typical SW bias seen in the all-day, multiyear mean (Figure 1) is also present during the JJA 2013 season in the northeast Pacific. Figure 2b shows the same top-of-atmosphere (TOA) upwelling SW radiation data separately for the model and the CERES observations along the MAGIC transect marked by diamonds in Figure 2a. Around longitude 137°W, the model's bias switches sign, with an overestimate of upwelling SW between 10 and 20 W/m^2 to the west and an underestimate of up to 75 W/m^2 near the Californian coast in the east.

Figure 3 shows an evaluation of all-sky LWP and LCC along the same transect against widely used data sets based on satellite microwave radiometer retrievals for LWP and the GOCCP CALIPSO lidar product for cloud

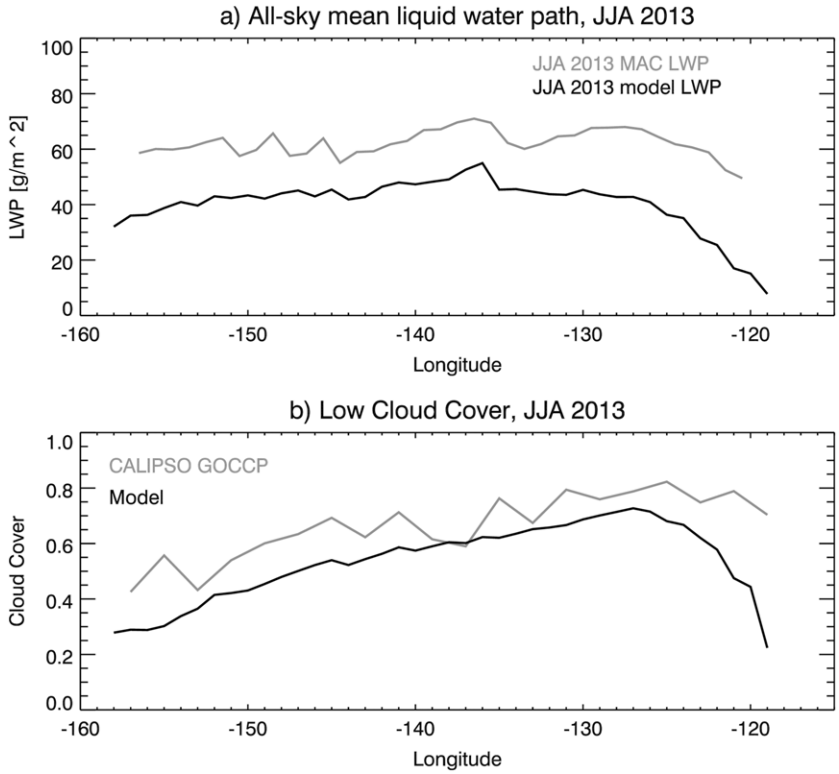

Figure 3. (a) All-day mean liquid water path from model and the MAC-LWP product, averaged for JJA 2013 along the MAGIC track. (b) All-day mean low cloud cover from the model and GOCCP CALIPSO product, averaged for JJA 2013 along the MAGIC track. MAGIC = Marine ARM GPCI Investigation of Clouds; ARM = Atmospheric Radiation Measurement; GPCI = GCSS Pacific Cross-section Intercomparison; GOCCP = GCM Oriented CALIPSO Cloud Product; LWP = liquid water path.

cover. As in Figures 1 and 2a, the fields shown here constitute averages across all times of day for the JJA 2013 season. The LWP retrieval is from the MAC-LWP product (Elsaesser et al., 2017), although results are very similar using the Wentz retrieval (Wentz, 2013) on data from individual satellites. The results suggest that the model underestimates both the LWP and the LCC all the way across the transect, concurrent with the *too few*, *too bright* picture in western half of the track, and a *too few*, *not bright enough* picture in the eastern half. This is consistent with the model's SW radiation bias in the stratocumulus region in the east, but inconsistent with the overestimated albedo in the more cumulus-dominated western part of the transect. This suggests that in the cumulus regime either cloud properties other than cover and LWP are in error, or there are significant ambiguities in the observational data and/or methodology of the comparison.

This provides the motivation for investigating differences between observational products as well as more carefully evaluating the colocation of model and observations in both space and time, and to consider properties of the cloud other than cover and LWP that may also affect the SW radiation, such as subgrid-scale cloud condensate heterogeneity, cloud liquid effective radius, cloud vertical overlap, and 3-D radiative effects. Given that the radiation bias under investigation is in the SW, it is of primary importance to assess the model during the daytime only (SW radiation present), and this is addressed first in the subsequent section, followed by an evaluation of cloud cover and LWP. Other aspects of the cloud and their radiative impacts are investigated in section 4.

3.1. Diurnal Cycle

As discussed above, SW biases are by definition limited to daytime conditions, yet cloud cover and water path observations often include measurements taken at different times of day or night (e.g., due to different local overpass times of polar-orbiting satellites) and are compared to a mean model state averaged over the entire 24 hr period. This is relevant for MBL clouds, which exhibit a distinct diurnal cycle, and so the following evaluation is performed with daytime data where possible.

The ship journey in the MAGIC campaign took around 5 days between ports thus sampling almost five diurnal cycles each way. The data from eight round trips are averaged together in order to get more robust statistics, which means that the diurnal cycle will be smoothed out in the composite. However, because the ship departed from port at similar times of day for each trip, the composite does show the dominant imprint of a diurnal cycle along longitude with maxima and minima in the solar irradiance, albeit smoothed due to the different directions of travel and slightly different track times. The diurnal cycle in the ship data is illustrated in Figure 2c; a composite of the surface solar irradiance (i.e., the downwelling component of the surface SW flux) for the eight MAGIC round trips shown together with model data matched in time and space. The shading in the background of the figure shows the density of samples at local ST and the black lines show the solar irradiance averaged across all transects. Considering that the TOA SW bias is on the order of 15 W/m^2 westward of 137°W, it is difficult to see a consistent sign of the bias on this partially time-smoothed background surface irradiance signal varying between 0 and 600 W/m^2. The model's larger overestimate of solar irradiance in the stratocumulus near the Californian coast is, on the other hand, clearly evident. The diurnal cycle in the MAGIC data is relevant for interpreting the model comparison in the following sections.

3.2. Cloud Cover

Cloud cover observations investigated are from the GOCCP lidar-based product, the CERES-MODIS SSF product, and the MAGIC ship ceilometer (lidar). The GOCCP climatology relies on observations obtained at around 1:30 a.m. and 1:30 p.m. ST, sampling different parts of the diurnal cycle of MBL cloudiness and providing separate all-day mean (i.e., the average of both overpass times, as shown in Figure 3) and daytime-only estimates. The daytime-only (1:30 p.m. ST) LCC from the GOCCP along the MAGIC track is shown in Figure 4a. Unfortunately, the single JJA 2013 season is not long enough to get a robust sample size from CALIPSO because of the narrow swath and large intertrack spacing of the data. The cloud cover fields are noisy for a single season, so the daytime LCC averaged over the JJA seasons from 2006 to 2016 is also included in Figure 4a to provide a more robust estimate of LCC. The instruments on board the Aqua satellite, on the other hand, have a greater viewing angle and thus a larger sample size to produce a more reliable estimate of daytime cloud cover for a single season. The CERES-MODIS SSF (Aqua) daytime cloud cover (local overpass time also approximately 1:30 p.m. ST) for the JJA 2013 season is similar to the 2006–2016 multiyear JJA daytime average but a few percent lower (Figure 4a). This figure illustrates that the 1:30 p.m. ST LCC averaged for the JJA season from 2006 to 2016 agrees remarkably well between GOCCP and CERES-MODIS SSF estimates, which gives confidence in the CERES-MODIS SSF estimate for the JJA 2013 season. Both satellite products may, however, underestimate LCC slightly if high clouds are obscuring the view of the boundary layer clouds. A second GOCCP product (Guzman et al., 2017) keeps track of the fraction of lidar profiles that fully attenuate (are opaque) as the lidar penetrates from TOA to the surface. Averaged over the JJA 2006–2016 seasons, the fraction of daytime profiles along the MAGIC transect that fully attenuate above the boundary layer (at altitude 3.12 km) is 2–3% along most of the track, with slightly higher values (up to 5%) at either end of the track (not shown). This fraction constitutes an upper limit to the low cloud potentially missed (if we assume every opaque profile hides a low cloud). Given that cloudiness is fractional, particularly in the western half of the track, a realistic underestimate of LCC from CALIPSO is therefore more likely on the order of a few percent and will not substantially change the conclusions drawn in this study.

Figure 4b compares the daytime cloud cover from CERES-MODIS SSF with the MAGIC ship data and the model. The LCC estimate from the MAGIC ceilometer includes any data with a retrieved cloud base below 6 km and is a composite for Legs 10–17. The diurnal cycle in the ceilometer-based composite is evident in the cloud cover, but the daytime minima along the composite track agree quite well with the daytime JJA 2013 CERES-MODIS SSF estimate from Aqua. When the model's LCC is averaged over JJA 2013 only for times around 1:30 p.m. ST to correspond with the Aqua satellite overpass time, the agreement with observations is good for much of the transect (Figure 4b), though the model does show a substantial underestimate of stratocumulus cloud cover close to the North American coast and to some degree near Hawaii. Given that the model's daytime cloud cover is in close agreement (difference <5%) with both the daytime-observed estimates in the trade cumulus area, it is unlikely to be the primary cause of the SW bias in this region.

So if the daytime cloud cover from the model agrees with the observations in the cumulus regime, the apparent lack of cloud cover versus the all-day mean GOCCP data shown in Figure 3b must be due to a nighttime underestimate. The two daily overpasses by CALIPSO cannot fully resolve the diurnal cycle, but since the overpass times correspond approximately to the maximum and minimum of cloudiness in MBL clouds, we use the difference between the 1:30 p.m. and 1:30 a.m. ST overpasses as a proxy for the diurnal range of LCC.

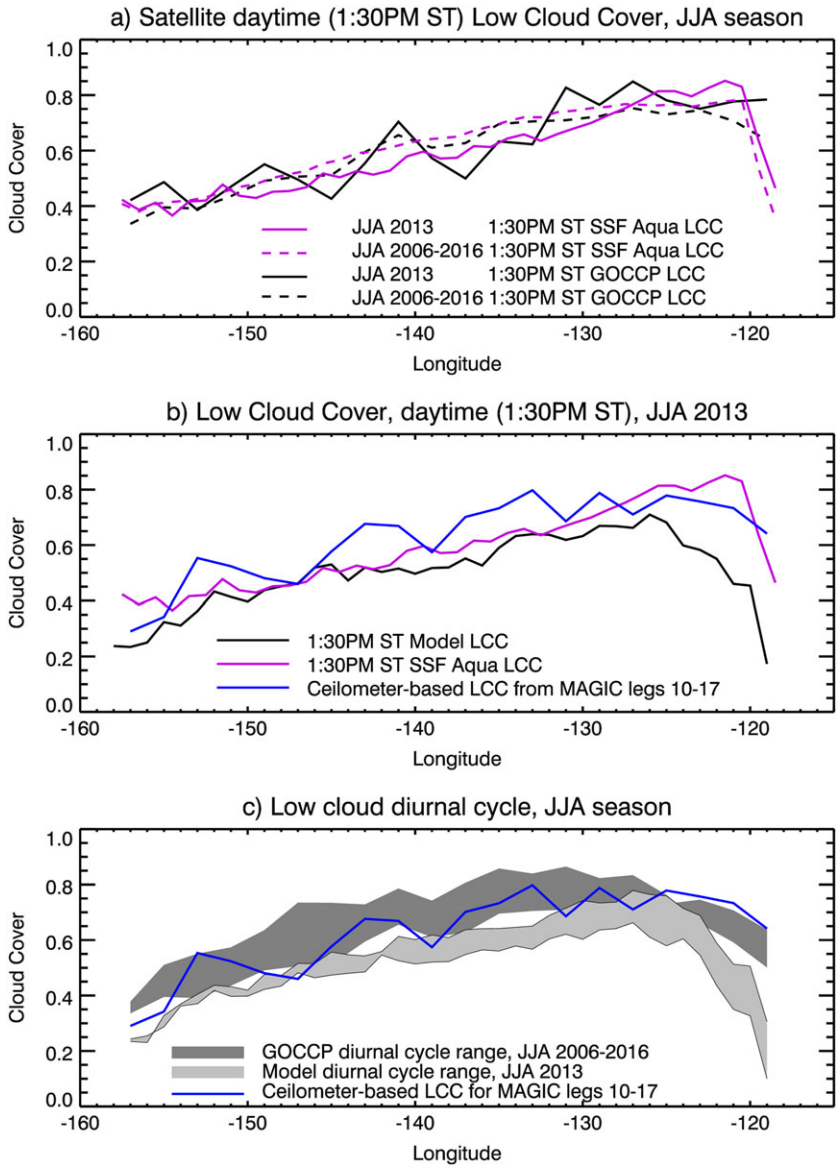

Figure 4. (a) Seasonal (JJA) mean LCC along the MAGIC track from CERES-MODIS SSF Aqua (purple) and GOCCP (black) products for the daytime (1:30 p.m. solar time, ST) overpass of the satellites, for single 2013 season (solid) and 10-year average 2006–2016 (dashed). (b) Daytime (1:30 p.m. ST) mean JJA 2013 CERES-MODIS SSF Aqua LCC (purple solid curve, as in panel a), model JJA 2013 LCC at time step closest to 1:30 p.m. ST (black) and composite of ceilometer-based low cloud cover from legs 10–17 of the MAGIC campaign (blue). (c) Diurnal range of LCC (i.e., range between cloud cover at 1:30 p.m. ST and 1:30 a.m. ST) from CALIPSO JJA 2006–2016 (dark gray shading), from the model for JJA 2013 (light gray shading), and composite ceilometer-based LCC from legs 10–17 of the MAGIC campaign (blue). JJA = June-July-August; LCC = low cloud cover; MAGIC = Marine ARM GPCI Investigation of Clouds; ARM = Atmospheric Radiation Measurement; GPCI = GCSS Pacific Cross-section Intercomparison.

The model's diurnal range in cloud cover (also calculated as the difference between 1:30 p.m. and 1:30 a.m. ST LCC) is compared to the observations in Figure 4c and shows the much smaller diurnal amplitude in the western half of the transect in the model compared to both the CALIPSO and the MAGIC ship data, explaining the discrepancy of about 5–10% cloud cover when comparing the all-day mean (Figure 3b).

There are always uncertainties in the observations, and there is a question of how reliable the amplitude of the diurnal cycle from CALIPSO is in the GOCCP product. A lower signal-to-noise ratio experienced by the lidar at nighttime might be suspected to contribute to an enhanced detection of cloud and therefore a larger apparent diurnal cycle. However, the backscatter signal from liquid boundary layer clouds is generally strong,

often attenuating the signal entirely within a few range bins, and it is unlikely that the lower signal-to-noise ratio during the day would result in missed cloud detection. The GOCCP algorithm also addresses the potential misidentification of aerosol as cloud at night by requiring a substantial signal-to-noise ratio of 5 for the identification of low liquid cloud, which should reliably exclude aerosol at day and night (Chepfer et al., 2010). The fact that the day/night span of the GOCCP data is also in reasonable agreement with the composite ceilometer-based cloud fraction from MAGIC (keeping in mind that the multiyear LCC in Figure 4c is a few percent higher than for the JJA 2013 season) also gives confidence in the results. Lastly, while the CERES-MODIS SSF product does not provide a nighttime-only cloud cover estimate, it does provide a combined day/night cloud cover, which is also in very good agreement with the day/night CALIPSO cloud cover (not shown), suggesting by inference that the nighttime agreement must be good as well.

3.3. Liquid Water Path

If daytime cloud cover is not the primary cause for the SW bias in the trade cumulus area, the next candidate is the amount of cloud condensate, that is, LWP. The availability of microwave radiances at daytime and nighttime from a number of satellite platforms makes them an attractive source of liquid water retrievals. However, while individual frequencies are more or less sensitive to various types of moisture, there remain difficulties in disentangling contributions from the sea surface, atmospheric vapor, cloud, and precipitation. This is particularly true in areas with broken cloud cover, such as in the trade cumulus regions. In this cloud regime, studies consistently find that LWP estimates from satellite MWR exceed those derived from visible wavelengths (CERES-MODIS; Lebsock & Su, 2014; Painemal et al., 2016; Seethala & Horváth, 2010; Zuidema & Joyce, 2008). Painemal et al. (2016) draw on the surface-based three-channel MWR from the MAGIC campaign as a third independent source of information and find that the MAGIC observations agree reasonably well with the CERES-MODIS SSF retrieval, but both provide lower LWP estimates than those from satellite MWR in trade cumulus. While the three-channel MWR retrieval is arguably the most accurate for nonprecipitating conditions with an uncertainty of about 5–8 g/m^2 (Fielding et al., 2015), a wet radome prevents accurate retrievals and will produce erroneously high LWP values. Thus, scenes where precipitation reaches the surface (which are likely to contain clouds with larger water path) must be removed, which will lead to a systematic underestimate of the all-sky average.

The CERES-MODIS SSF retrieval is also subject to uncertainties, and larger water paths could be retrieved if alternative, reasonable assumptions were made in the algorithm. Retrievals from various MODIS-visible wavelengths, for example, represent effective radii detected at different penetration depths into the cloud top (Platnick, 2000), and the 3.7-μm wavelength chosen here results in systematically smaller effective radii (and therefore lower LWP estimates) than those from the 1.6- or 2.1-μm channels (Zhang et al., 2017). On the other hand, the 3.7-μm channel is less affected by precipitation, arguably making it the best choice for estimating the cloud-only water path (Lebsock et al., 2011). Alternatively, the current assumption that the cloud water profile is adiabatic could be relaxed to allow a more uniform (or homogeneous) cloud water profile, which would also result in an increased LWP of up to 20% (Bennartz, 2007).

Lebsock and Su (2014) draw on additional retrievals from CloudSat and CALIPSO to investigate the cause of the discrepancies between microwave and visible LWP retrievals, and their results suggest that the greatest contribution to the microwave overestimate of LWP (in the trade cumulus regime) may come from misidentifying precipitation in the column as cloud water. Smaller contributions can be explained by errors in the clear-sky detection and solar zenith angle dependency of MODIS observations, and a clear-sky bias in the MWR retrieval. While the Lebsock and Su (2014) study uses an alternative CERES-MODIS retrieval (Collection 5.1; Platnick et al., 2003) to the one used here and in Painemal et al. (2016), the conclusions are likely to apply more broadly, given that both CERES-MODIS retrievals produce comparable results in MBL clouds (Zhang et al., 2017). Greenwald et al. (2018) show that if a bias correction is applied to the satellite MWR retrieval accounting for precipitation-related biases (among others), the LWP retrieved from satellite MWR is reduced and becomes more comparable to the lower water paths from MAGIC in broken cloud conditions.

Figure 5a shows the JJA 2013 (Terra and Aqua) CERES-MODIS SSF all-sky LWP retrieval together with the model all-sky LWP for solar times to match the Terra/Aqua overpass at 10:30 a.m. and 1:30 p.m., respectively. For the satellite product, estimates assuming adiabatic and homogeneous condensate profiles are shown. Relative to both of these, the model overestimates all-sky LWP in the cumulus regime and underestimates it in stratocumulus regime, though the point of bias reversal shifts further west if comparing against the retrieval assuming

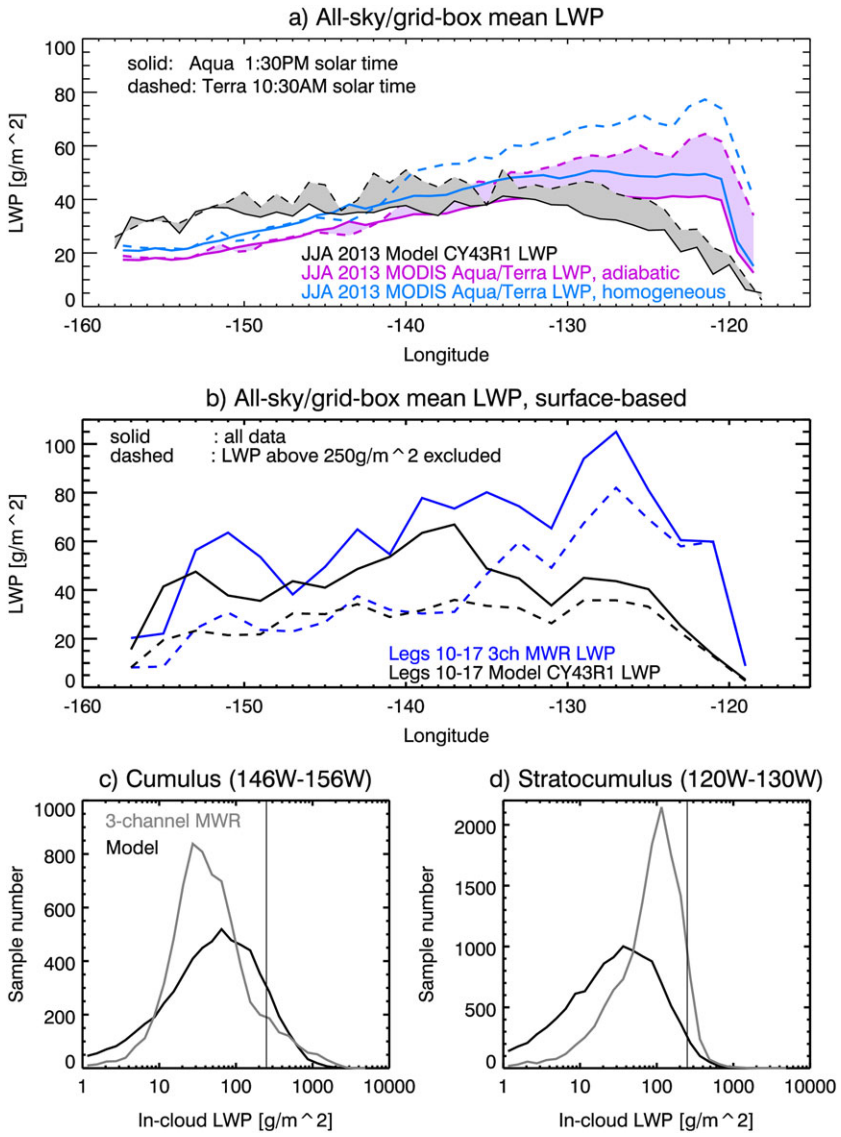

Figure 5. (a) All-sky LWP estimates averaged for the JJA 2013 season for (solid) Aqua and (dashed) Terra overpass times from the CERES-MODIS SSF retrieval assuming (purple) an adiabatic or (blue) a homogeneous cloud condensate profile, and (black) from the IFS. (b) Composite of (blue) three-channel MWR retrieved all-sky LWP from MAGIC legs 10–17 and (black) matched IFS model data. Solid curves show the all-sky LWP if all samples are included (including those affected by precipitation), while dashed curves show the all-sky LWP estimate if in-cloud LWP samples exceeding 250 g/m^2 are treated as missing data. (c) Distribution of in-cloud LWP samples observed by the (gray) three-channel MWR for the cumulus regime, and (black) calculated from the IFS using assumptions consistent with those used in the radiation scheme. (d) Same as (c) but for the stratocumulus regime. Model sample numbers have been scaled to be comparable to observation sample numbers. LWP = liquid water path; JJA = June-July-August; MAGIC = Marine ARM GPCI Investigation of Clouds; ARM = Atmospheric Radiation Measurement; GPCI = GCSS Pacific Cross-section Intercomparison; LCC = low cloud cover.

homogeneous profiles. Qualitatively, the bias relative to the adiabatic estimate, which switches signs around 137°W, fits with the sign of the observed SW bias (Figure 2b).

Figure 5b shows the MAGIC three-channel MWR all-sky LWP composite across legs 10–17 (including observations taken at all times of day) and compares these to the model data matched in time and space to the ship's track. The calculation of the all-sky LWP is highly sensitive to how precipitation-contaminated samples are treated. If all samples are included (solid curves), the model all-sky LWP significantly underestimates the observed LWP in the eastern half of the track, but the comparison is more ambiguous in the western half. However, this is not a good comparison, given that under precipitating conditions the MWR retrieval has

large errors. If instead all in-cloud LWP samples from both the observations and the model above a (somewhat arbitrary) threshold of 250 g/m^2 are excluded (dashed), the observed all-sky LWP is reduced significantly. This appears to improve the agreement in the western half of the track, but LWP in the eastern half remains underestimated by the model.

The comparison of the model with the three observational estimates gives a model all-sky LWP in the cumulus regime that is too low relative to satellite MWR retrievals (MAC-LWP in Figure 3a), too high relative to the CERES-MODIS SSF retrieval (Figure 5a), and apparently comparable to the surface-based MWR observations (Figure 5b). In the stratocumulus regime, the message is more consistent and suggests that the model underestimates all-sky LWP. While some of these discrepancies, particularly the relative overestimate of LWP from satellite MWR, can be at least partly explained by biases related to precipitation, the threshold sensitivity of the surface-based observations suggests that a low number of high-value LWP samples disproportionally affect the all-sky average. This calls into question whether the all-sky average is the most appropriate measure of condensate amount to compare if the goal is to explain the SW bias, especially as the SW albedo is more sensitive to intermediate water path values and saturates for very high water paths (Cahalan et al., 1994).

A more instructive way to evaluate water path agreement may therefore be to consider the full frequency distributions of instantaneous in-cloud LWP, as it is the midrange in-cloud values of the condensate amount that are most relevant for the SW radiative transfer. Figures 5c and 5d show in-cloud LWP frequency distributions from the MAGIC three-channel MWR for two longitude ranges representative of the cumulus regime (146°W to 156°W) and the stratocumulus regime (120°W to 130°W), alongside the equivalent distributions of in-cloud LWP from the IFS, matched in time and space to the MAGIC observations. The in-cloud LWP values for the model were calculated using assumptions consistent with those used in the IFS radiation calculations, as outlined below.

The IFS radiation code is based on the Rapid Radiation Transfer Model (RRTM; Iacono et al., 2008; Mlawer et al., 1997). Cloud-radiation interactions are taken into account using the Monte Carlo Independent Column Approximation method where the subgrid cloud properties (cloud fraction, vertical overlap, and in-cloud condensate heterogeneity) are represented by splitting each grid column into a number of subcolumns for the radiation calculations (McRad; Morcrette et al., 2008). The grid box condensate amount is distributed across 112 subcolumns (for the SW radiative transfer) assuming that the in-cloud condensate is drawn from a Gamma distribution whose mean corresponds to the grid box average in-cloud condensate amount and whose width is chosen such that the standard deviation of the distribution equals the mean (i.e., the fractional standard deviation, FSD, which is the standard deviation divided by the mean, is equal to 1). Exponential overlap is then applied to align condensate amount on subcolumns in the vertical. The resulting subcolumn LWP is the model's closest representation to the near-instantaneous in-cloud LWP observed by the three-channel MWR and is the water path used for the radiation transfer calculations.

It is evident from Figures 5c and 5d that the model's mode of the in-cloud LWP frequency distribution is shifted to higher values in the cumulus regime and shifted to lower values in the stratocumulus regime. The high-end tail of the distribution remains highly uncertain from the observations, but in fact, matching the mode of the distribution is more important for the SW than the tail. The figure also illustrates why merely excluding samples above 250 g/m^2 (or a similar threshold) does not make the all-sky LWP comparison more informative, as different proportions of the distributions are removed in observations and model, without affecting the mode mismatch. In order to improve the clouds' SW albedo, the agreement of the in-cloud LWP distributions must be improved.

A shift of the distribution mode can be achieved by reducing the model's bulk condensate amount in the cumulus regime and increasing it in the stratocumulus regime, a change that is consistent with adjusting the model toward observed all-sky LWP values from the CERES-MODIS SSF retrieval. Given the recent insights from Lebsock and Su (2014) and Greenwald et al. (2018) on precipitation-related biases from the satellite MWR retrievals, the CERES-MODIS SSF LWP therefore seems to the be the more appropriate reference choice here, despite the remaining uncertainties on the magnitude of the all-sky LWP.

4. Off-Line Radiative Transfer Experiments

We run an off-line version of the IFS radiation scheme (ECRAD; Hogan & Bozzo, 2016) on the model output along the MAGIC transect to test how changes to the model condensate amount and cloud cover impact the

in-cloud LWP distribution and SW bias. The off-line radiation code is run on 3-hourly snapshots of the model state along the track for the JJA 2013 period. The same model hindcasts as described in section 2 are used, and tracks are extracted for the 12- to 33-hr forecast range. In order to more fully sample the diurnal cycle, the radiation calculations are performed three times on each 3-hourly model state, adjusting the solar zenith angle to the preceding and following hour, giving an hourly time series along the ship track. Averaged over the JJA 2013 period, this off-line radiation calculation closely reproduces the online TOA SW bias seen in the model hindcasts discussed earlier (Figure 2b). The impact of adjusting different aspects of the cloud on the radiation can then be investigated by modifying the model output to agree more closely with the observations. Incrementally correcting the model's bulk condensate and cloud cover parameters to agree with observed values impacts the mode and shape of the in-cloud LWP distribution. This is followed by an evaluation of other potential sources of error, including the effective radius, assumptions about subgrid-scale horizontal cloud heterogeneity, vertical overlap, and 3-D radiative effects.

4.1. Liquid Water Path

As the model's daytime cloud cover is quite realistic in the cumulus regime, the most straightforward way to substantially shift the mode of the in-cloud LWP distribution toward lower values is to reduce the model's condensate amount, a measure of which is the all-sky LWP. In the stratocumulus regime, all observational data sets agree that the all-sky LWP is underestimated, and an increase would lead to a more realistic in-cloud LWP distribution, regardless of the coincident cloud cover bias. For the first off-line experiment, we therefore adjust the model's grid box mean condensate amount until the average JJA 2013 all-sky LWP along the transect agrees with the CERES-MODIS SSF retrieval (assuming adiabatic condensate profile). Since Terra's overpass time (10:30 ST) samples LWP approximately halfway between the diurnal minimum and maximum (and therefore closer to the 24-hr average), the Terra JJA 2013 LWP is used as a reference. The grid box mean cloud liquid water content is enhanced or reduced along the tracks by a factor (fixed in time and height, but varying in longitude) at each grid point such that the JJA 2013 model LWP agrees with the Terra LWP along the MAGIC transect (Figure 6a). This shifts the mode of the in-cloud LWP distribution close to the observed value in the cumulus region in the west (Figure 7b, purple curve) and partway toward the observed mode in the eastern stratocumulus region (Figure 7c, purple curve). The TOA upwelling SW bias all but disappears in the western half of the track (Figure 7a, purple curve). In the eastern half of the track, the SW error is reduced, but a significant bias remains near the Californian coast, likely reflecting the model's underestimate in cloud cover in the stratocumulus regime.

4.2. Cloud Cover

Combining the modified LWP with a modified cloud cover (multiplying the model cloud fraction at each grid point along the track by a factor, fixed in time and height, by varying in longitude, to approximate the observed Terra LCC), the TOA SW bias is further reduced in the stratocumulus regions east of 130°W (Figure 7a, blue curve). As expected, there is little impact west of 130°W because the model cloud cover is already close to the observations, but there is a small improvement close to Hawaii where the model underestimated the observed LCC. About half of the bias remains in the stratocumulus region in the east. A look at the in-cloud LWP distribution (Figure 7c, blue curve) shows that the mode of the distribution remains underestimated, and the width is greater than observed due to the increase in cloud cover reducing the in-cloud LWP. It is possible that the adjusted cloud cover is still imperfect, as a lack of cloud occurrence cannot be fully compensated for by increasing cloud cover when present. However, there are also other sources of potential bias investigated below.

4.3. Effective Radius

The effective radius of liquid water cloud is used in the radiation calculation of the optical depth and is parameterized in the IFS according to Martin et al. (1994) with an enhancement of the effective radius for drizzling cloud based on Wood (2000). The CDNC is calculated from a combination of a background value assumed over ocean, modified slightly by wind speed (Erickson et al., 1986; Genthon, 1992). As the wind speed along the MAGIC transect does not change systematically with longitude, the CDNC from the model is fairly constant with a value of around 50 cm^{-3} along the entire track (Figure 8a). The effective radius is also relatively constant along the track with an average value of 10 μm, with greater values (12–13 μm) near cloud top and lower values (7–8 μm) near cloud base due to higher cloud water content at the cloud top (Figure 8b).

Alongside the model values, Figure 8 also shows the CDNC concentration derived from the MAGIC UHSAS observations by McGibbon and Bretherton (2017) and averaged for the eight MAGIC round trips during JJA

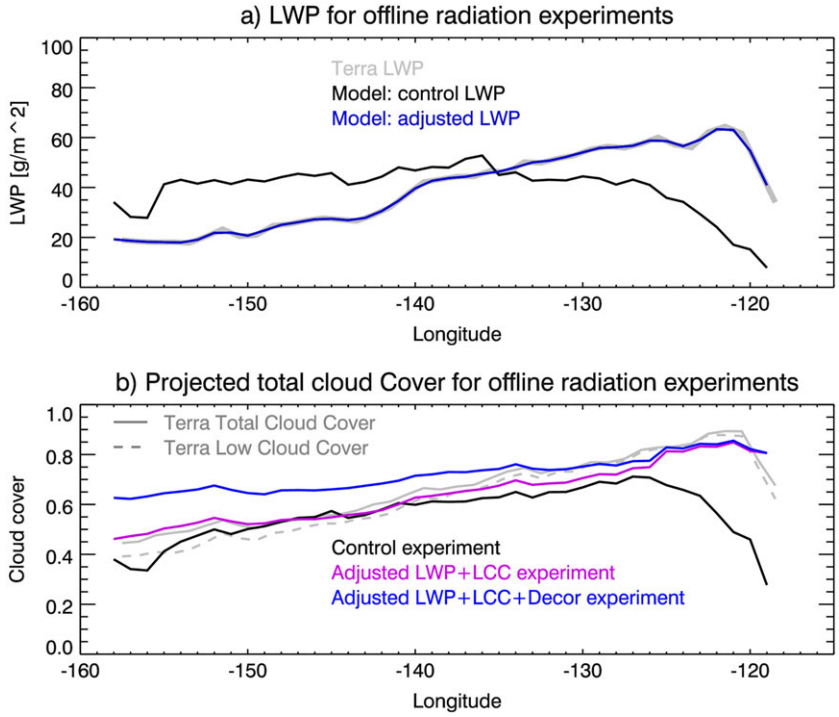

Figure 6. (a) JJA 2013 mean LWP for (black) control off-line model state, (blue) adjusted model LWP to agree with (gray) JJA 2013 mean Terra LWP. (b) Projected total cloud cover from off-line radiation experiments. (gray) Terra observed low and total cloud cover, (black) projected total cloud cover determined by off-line radiation scheme for control model state, (purple) from experiment with adjusted LWP and cloud cover to better agree with Terra observed state, (blue) from experiment with adjusted LWP, cloud cover, and reduced overlap decorrelation length scale.
JJA = June-July-August; LWP = liquid water path; LCC = low cloud cover.

2013 (Figure 8a), and the CERES-MODIS SSF effective radius (from the VNIR and 3.7-μm channels, consistent with the LWP product) averaged for JJA 2013 from the Aqua and Terra daytime retrievals (Figure 8b). These observation-derived parameters suggest a gradient from more polluted conditions and smaller cloud drop effective radius near the North American coast, to cleaner conditions and larger cloud droplets over the open ocean.

There remain uncertainties related to both the CDNC derivation from the UHSAS and the effective radius retrieval. The CERES-MODIS SSF product has been shown to overestimate in situ measured effective radius in southeast Pacific MBL clouds by about 15–20% for the VNIR and 2.1-μm channel combination, with slightly lower values (and thus a reduced overestimate) found for the VNIR and 3.7-μm channel retrieval (Painemal & Zuidema, 2011). Zhang et al. (2017) also find that the MODIS effective radius at the 3.7-μm wavelength is about 10% greater than that derived from surface-based retrievals of MBL clouds in the Azores. It is therefore likely that the effective radius for the VNIR and 3.7-μm channel combination is an overestimate by around 10%.

For this off-line radiation experiment, we choose to replace the default CDNC in the model with the JJA 2013 average values along the MAGIC transect. The resulting effective radius at cloud top is plotted as the blue curve in Figure 8b. While this effective radius does not agree with the CERES-MODIS SSF observations along the entire track (underestimating values at either end), it does capture the gradient from coast to Hawaii and is self-consistent with the model's cloud cover and water content.

Figure 7a (green curve) shows the TOA upwelling SW when this new effective radius is used (in addition to the previously adjusted LWP and LCC) in the off-line radiation calculations. As expected, given the change in effective radius relative to the default model, the cloud albedo is significantly enhanced (up to 15 W/m^2) toward the Californian coast. This constitutes a further improvement in the SW bias in the stratocumulus region. The increase in effective radius in the western half of the track has less impact on the SW (a 2- to 3-W/m^2 reduction in upwelling SW), in part because the relative change in effective radius is smaller but also due to the smaller cloud fractions encountered here.

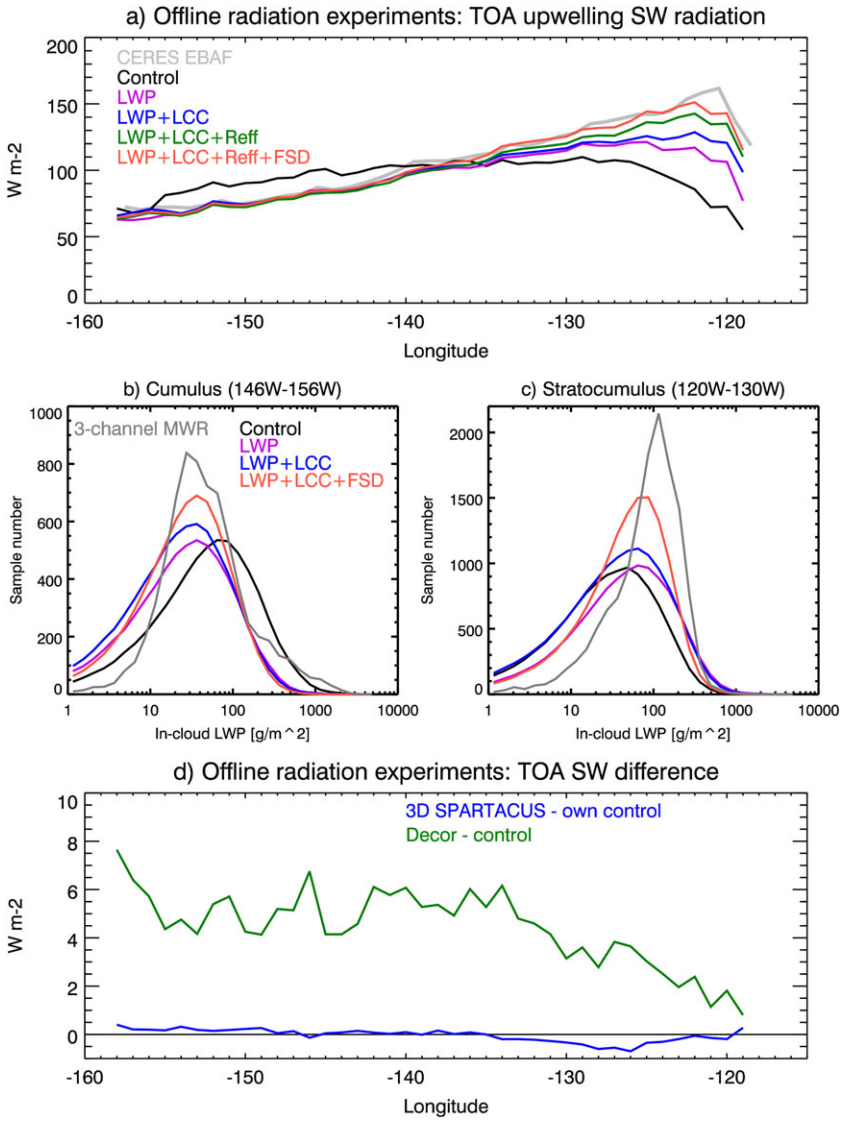

Figure 7. Result of off-line radiation calculations (a) TOA upwelling SW radiation for the JJA 2013 season from (gray) CERES EBAF, (black solid) the control radiation scheme, (purple) with adjusted LWP, (blue) with adjusted LWP and cloud cover, (green) with adjusted LWP, cloud cover, and using the UHSAS observations to derive effective radius, and (red) with adjusted LWP, cloud cover, effective radius, and using the regime-dependent cloud heterogeneity parameterization. (b, c) In-cloud LWP distributions for the cumulus and stratocumulus regimes from three-channel MWR (gray) and from JJA 2013 3-hourly model data along the MAGIC track used for off-line radiation experiments. (d) TOA upwelling SW radiation differences between (blue) the SPARTACUS radiation solver including a representation of 3-D radiative effects, and its own control (Tripleclouds), and (green) using a reduced decorrelation length scale for overlap versus control. TOA = top-of-atmosphere; SW = shortwave; JJA = June-July-August; LWP = liquid water path; LCC = low cloud cover; FSD = fractional standard deviation; MAGIC = Marine ARM GPCI Investigation of Clouds; ARM = Atmospheric Radiation Measurement; GPCI = GCSS Pacific Cross-section Intercomparison.

4.4. Subgrid-Scale Cloud Heterogeneity

Lebsock et al. (2013), Boutle et al. (2014), and later Ahlgrimm and Forbes (2016) show evidence that cloud heterogeneity is greater in the trade cumulus regime than in the stratocumulus regime and suggest a regime-dependent parameterization to diagnose the FSD (the standard deviation divided by the mean) of cloud liquid. As described in section 3.3, this parameter can then be used in the radiation scheme of the IFS to prescribe the in-cloud condensate amount on subcolumns, which are generated as part of the Monte Carlo Independent Column Approximation approach. The current default value for FSD is 1, while the parameterization developed by Ahlgrimm and Forbes (2016) suggests values around 0.5 in the stratocumulus regions and around 0.9 in the cumulus regime; that is, in-cloud condensate in stratocumulus is more homogeneous

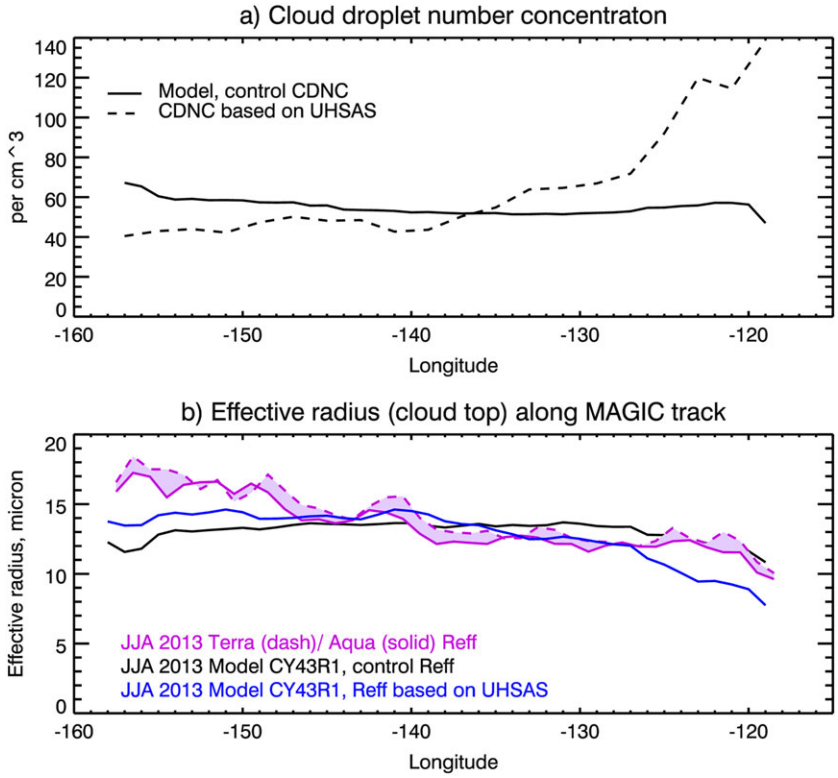

Figure 8. (a) Estimates of CDNC concentration along MAGIC track, derived from (dashed) UHSAS observations aboard MAGIC from Legs 10–17, and (black) from the default parameterization in the IFS. (b) Model effective radius calculated off-line based on 3-hourly model data extracted along MAGIC transect during JJA 2013 using (black) the default scheme used in the IFS and (blue) using the UHSAS-derived CDNC. The effective radius from CERES-MODIS SSF (VNIR and 3.7-μm channels) from Aqua and Terra averaged for JJA 2013 is shown in purple. CDNC = cloud droplet number concentration; MAGIC = Marine ARM GPCI Investigation of Clouds; ARM = Atmospheric Radiation Measurement; GPCI = GCSS Pacific Cross-section Intercomparison; IFS = Integrated Forecast System; UHSAS = Ultra-High Sensitivity Aerosol Spectrometer.

than for cumulus. Based on the remaining disagreement of the in-cloud LWP distributions, a lower FSD (and therefore narrower distribution) would be expected to improve the fit. Implementing this parameterization leads to an enhanced cloud albedo overall, since the current fixed FSD value is higher than that supported by observations. However, the enhancement of the cloud albedo is differentially greater in the stratocumulus region than in the cumulus regime. In Figure 7a, the impact of the parameterization is shown in addition to the previously discussed adjustments to LWP, LCC, and effective radius (red curve). As expected, the impact is greatest in the stratocumulus regime, increasing the upwelling SW by another 8 W/m^2 and bringing the SW in very close agreement with the observations. The impact is small along the western half of the track, where the parameterized FSD value remains similar to the previously assumed fixed value of 1. The adjusted FSD does not change the mode of the in-cloud LWP distribution (Figures 7b and 7c, red curve) but does sharpen the distribution giving closer agreement with the observed distribution.

4.5. Cloud Vertical Overlap

The radiation scheme operational in the IFS CY43R1 assumes an exponential cloud overlap (Barker, 2008) with a latitude-dependent decorrelation length varying between 2.2 and 2.6 km along the MAGIC transect, based on Shonk et al. (2010). The literature provides a variety of estimates of a decorrelation length scale of order of a few kilometers (Hogan & Illingworth, 2000; Mace & Benson-Troth, 2002; Oreopoulos & Khairoutdinov, 2003; Tompkins & Di Giuseppe, 2015) but Neggers et al. (2011) argue, based on LES simulations, that the decorrelation length scale in boundary layer clouds should actually be much smaller, around 220 m. As a sensitivity test we reduce the decorrelation length scale in the off-line radiation calculations by a factor of 10, giving a value close to that suggested by Neggers et al. (2011). The reduced decorrelation length scale increases the projected total cloud cover toward the west and by up to 15% near Hawaii (Figure 6b). As one might expect, the impact is greatest for broken cloud cover, adding about 4.5 W/m^2 to the upwelling SW radiation at the

TOA in the trade cumulus region, a bit less near the Californian coast (Figure 7d). This change is not included in Figure 7a as the net effect of the overlap change translates into a greater projected cloud cover, and cloud cover was already optimized in a previous step. However, it illustrates the potential impact on the SW radiation of uncertainties in assumptions in the parameterization of cloud in the model.

4.6. Three-Dimensional Radiative Effects

Another uncertainty in the impact of cloud on radiation is the importance of representing 3-D effects compared to the 1-D plane-parallel assumption used in most global models. The SPeedy Algorithm for Radiative TrAnsfer through CloUd Sides (SPARTACUS; Hogan et al., 2016; Schäfer et al., 2016) includes a treatment of 3-D radiative effects. The option to switch from the McRad solver to the SPARTACUS solver is part of the new ECRAD package, and we use it here to quantify the impact of 3-D radiative effects. The Tripleclouds solver (Shonk & Hogan, 2008) is also part of the ECRAD package and is the appropriate control for showing the impact of the 3-D effects from SPARTACUS (as shown in Figure 7d). Depending on the solar zenith angle and cloud geometry (captured by a parameter that estimates the representative cloud size), including the 3-D effects leads to less SW radiation reaching the surface when the Sun is low, as the cloud sides intercept radiation, while at high angles more solar radiation reaches the surface, from trapping of solar radiation due to horizontal transfer beneath the highest cloud layer. As expected, the inclusion of 3-D radiation acts to reduce or enhance the cloud radiative effect at different times of day, based on the solar zenith angle. In absolute terms, the impact in the particular cloud regimes considered here is very moderate when averaged over a 24-hr period, enhancing the upwelling SW along the MAGIC transect by about 0.5 W/m^2 in the western half of the track, while reducing it by up to 0.8 W/m^2 between 120°W and 140°W (Figure 7d, blue curve). As the 3-D effect along the transect is small compared to the impact of other changes, it does not significantly change the SW bias and is not included in Figure 7a for clarity.

5. Discussion and Conclusions

The aim of this paper is to improve our understanding of the possible causes of systematic SW radiation errors in subtropical MBL cloud regimes, errors that are common to many global models used for weather forecasting and climate projections. A range of observations from different satellite instruments and surface-based observations from the ARM MAGIC ship campaign in the northeast Pacific are used to quantify systematic errors in cloud properties in a global model, the ECMWF IFS. The comparison of observations and model is performed along the ship transect crossing both boundary layer cloud regimes of stratocumulus and cumulus that have opposite signs in the SW bias. The effects of the different cloud errors on the SW bias are then evaluated in off-line radiation calculations and, when combined, are able to explain the total SW bias in both cloud regimes.

As might be expected, systematic errors in the SW radiation are most sensitive to biases in cloud cover and LWP. When using model output and observations averaged across all times of day the IFS appears to underestimate LWP and cloud cover in the cumulus regime—biases of the opposite sign to explain the too high albedo for this cloud regime. There are two reasons for this misinterpretation of the evaluation. First, because the model underestimates the observed diurnal cycle of cloud cover and water path in both cloud regimes, the systematic errors in the daytime (relevant for the SW bias) are different to the errors in the 24-hr mean. In the cumulus regime, there is good agreement of cloud cover in the daytime, but the model does not capture the nighttime increase in cloud cover adequately. This results in a low bias in the 24-hr mean, which nevertheless does not have a negative impact on the SW radiation. In contrast, for the stratocumulus regime daytime and nighttime cloud cover is underestimated, and this does contribute to the SW bias.

Second, in the broken cloud regime, there are significant differences between the LWP products based on satellite microwave radiometer retrievals, the ship-based microwave radiometer, and the CERES-MODIS SSF satellite retrieval. The differences between observational data sets change the apparent sign of the model LWP bias in this region. While it remains challenging to measure and constrain the all-sky LWP conclusively, evidence from recent studies suggests that the LWP retrieved from satellite MWR in this cloud regime is likely an overestimate that can be explained in part by retrieval biases related to the presence of precipitation. Analysis of the full in-cloud LWP frequency distribution from the surface-based MWR retrieval also suggests that the all-sky LWP may not be the most suitable measure to consider, given that this average is sensitive to the small number of samples in the high-end tail of the distribution. The SW albedo response saturates for large water paths, so errors in the high-end tail will have little impact on the SW bias. Compared to these

observed distributions, the mode of the model's in-cloud LWP distribution is shifted toward higher LWP values in the cumulus regime, consistent with the too bright cloud scenario. In the stratocumulus regime a shift of the mode in the opposite direction is found, which is also consistent with the sign of the SW bias. While several parameters control the shape and mode of the model's in-cloud LWP distribution, the position of the mode can be improved by adjusting the model's all-sky LWP toward the CERES-MODIS SSF estimate.

A series of sensitivity tests using off-line radiation calculations demonstrate that making the characteristics of the model's boundary layer clouds consistent with the observations can largely explain the SW radiation bias across the cloud regimes. For the trade cumulus regime in the western half of the domain, given that the IFS cloud cover and subgrid cloud variability are already close to observed, it is a reduction of the all-sky LWP that has the biggest impact on shifting the mode of the in-cloud LWP distribution and thus the SW bias. For the stratocumulus-dominated regime in the eastern half of the domain, it is a combination of deficiencies in the model cloud that explains the SW bias. Increasing both the cloud cover and the LWP closer to observations, as well as improving the effective radius and reducing the subgrid heterogeneity of cloud condensate in overcast stratocumulus, can greatly reduce the SW bias in this regime. Three-dimensional radiative effects of the cloud are shown to be of less importance for the situation investigated here (MBL clouds and considering SW radiation averaged across all times of day).

There are clearly uncertainties in all aspects of the evaluation, and the match between modeled and observed in-cloud LWP distributions, while improved, remains imperfect. This could indicate that some compensating errors still persist. It is possible that by focusing on the SW radiation and adjusting the model to agree with the lowest of the observed all-sky LWP estimates, we now underestimate the number of high-LWP samples, which could not be reliably constrained by the surface-based observations. In other applications, for example, considering autoconversion and accretion processes, or indeed for forward modeling microwave radiances, the high-end tail of the LWP distribution will be more important. It is maybe also not entirely surprising that the adjusted model should reproduce observed radiative fluxes closely, given that the CERES-MODIS SSF LWP and LCC used as a reference in this study are derived to be consistent with the observed CERES TOA radiation. However, not all elements relevant to the model's radiative transfer were chosen based on this retrieval (effective radius, heterogeneity parameterization from independent sources), leaving room for disagreement.

Uncertainties with impacts smaller than around 5 W/m^2 may be difficult to constrain, which in any case is close to the accuracy of the CERES SW observations (Loeb et al., 2018). While the relative contributions from various sources to the model bias are specific to the IFS, a similar analysis using the appropriate observations should be equally enlightening when applied to other models.

This analysis also illustrates the value of exploring model biases in the numerical weather prediction context where the model state remains close to observations, allowing a like-for-like comparison. With a deeper understanding of the model errors gained through this approach, informed decision can be made to continue the improvement of cloud representation in global models.

Acknowledgments
This work was supported by U.S. Department of Energy's Office of Science via the Atmospheric System Research program under grant DE-SC0005259. Data used in this study are accessible via the URLs and DOIs listed in Table 1. Particular thanks to Maria Cadeddu for her expertise in producing and interpreting the three-channel MWR retrieval from MAGIC and to David Painemal for advice and interpretation of the CERES-MODIS SSF satellite products. We are grateful to Jeremy McGibbon and Chris Bretherton for sharing their UHSAS-derived CCN data set with us. We appreciate the contributions from our colleagues at ECMWF, to this paper through many fruitful discussions, including Peter Bechtold, Alan Geer, and Katrin Lonitz for providing valuable comments. We would like to thank Matt Lebsock and two anonymous reviewers for their critical reviews and constructive suggestions that have resulted in significant improvements to the paper.

References

Ahlgrimm, M., & Forbes, R. M. (2016). Regime dependence of cloud condensate variability observed at the atmospheric radiation measurement sites. *Quarterly Journal of the Royal Meteorological Society*, *142*(697), 1605–1617.

Barker, H. W. (2008). Representing cloud overlap with an effective decorrelation length: An assessment using cloudsat and CALIPSO data. *Journal of Geophysical Research*, *113*, D24205. https://doi.org/10.1029/2008JD010391

Bennartz, R. (2007). Global assessment of marine boundary layer cloud droplet number concentration from satellite. *Journal of Geophysical Research*, *112*, D02201. https://doi.org/10.1029/2006JD007547

Bony, S., & Dufresne, J.-L. (2005). Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophysical Research Letters*, *32*, L20806. https://doi.org/10.1029/2005GL023851

Boutle, I., Abel, S., Hill, P., & Morcrette, C. (2014). Spatial variability of liquid cloud and rain: Observations and microphysical effects. *Quarterly Journal of the Royal Meteorological Society*, *140*(679), 583–594.

Cadeddu, M., Liljegren, J., & Turner, D. (2013). The Atmospheric Radiation Measurement (ARM) program network of microwave radiometers: Instrumentation, data, and retrievals. *Atmospheric Measurement Techniques*, *6*(9), 2359.

Cahalan, R. F., Ridgway, W., Wiscombe, W. J., Bell, T. L., & Snider, J. B. (1994). The albedo of fractal stratocumulus clouds. *Journal of the Atmospheric Sciences*, *51*(16), 2434–2455.

Chepfer, H., Bony, S., Winker, D., Cesana, G., Dufresne, J., Minnis, P., et al. (2010). The GCM-oriented calipso cloud product (CALIPSO-GOCCP). *Journal of Geophysical Research*, *115*, D00H16. https://doi.org/10.1029/2009JD012251

Elsaesser, G. S., O'Dell, C. W., Lebsock, M. D., Bennartz, R., Greenwald, T. J., & Wentz, F. J. (2017). The multisensor advanced climatology of liquid water path (MAC-LWP). *Journal of Climate*, *30*(24), 10,193–10,210.

Elsaesser, G. S., O'Dell, C. W., Lebsock, M. D., & Teixeira, J. (2016). Multisensor advanced climatology mean liquid water path L3 monthly 1 degree × 1 degree V1. https://doi.org/10.5067/MEASURES/MACLWPM

Erickson, D., Merrill, J., & Duce, R. (1986). Seasonal estimates of global atmospheric sea-salt distributions. *Journal of Geophysical Research*, *91*(D1), 1067–1072. https://doi.org/10.1029/JD091iD01p01067

Fielding, M., Chiu, J., Hogan, R., Feingold, G., Eloranta, E., O'Connor, E., & Cadeddu, M. (2015). Joint retrievals of cloud and drizzle in marine boundary layer clouds using ground-based radar, lidar and zenith radiances. *Atmospheric Measurement Techniques*, *8*(7), 2663.

Genthon, C. (1992). Simulations of desert dust and sea-salt aerosols in Antarctica with a general circulation model of the atmosphere. *Tellus B*, *44*(4), 371–389.

Greenwald, T. J. (2009). A 2 year comparison of AMSR-E and MODIS cloud liquid water path observations. *Geophysical Research Letters*, *36*, L20805. https://doi.org/10.1029/2009GL040394

Greenwald, T. J., Bennartz, R., Lebsock, M., & Teixeira, J. (2018). An uncertainty data set for passive microwave satellite observations of warm cloud liquid water path. *Journal of Geophysical Research: Atmospheres*, *123*, 3668–3687. https://doi.org/10.1002/2017JD027638

Guzman, R., Chepfer, H., Noel, V., Vaillant de Guélis, T., Kay, J., Raberanto, P., et al. (2017). Direct atmosphere opacity observations from CALIPSO provide new constraints on cloud-radiation interactions. *Journal of Geophysical Research: Atmospheres*, *122*, 1066–1085. https://doi.org/10.1002/2016JD025946

Hilburn, K., & Wentz, F. (2008). Intercalibrated passive microwave rain products from the unified microwave ocean retrieval algorithm (UMORA). *Journal of Applied Meteorology and Climatology*, *47*(3), 778–794.

Hogan, R. J., & Bozzo, A. (2016). *ECRAD: A new radiation scheme for the IFS*. Englang: European Centre for Medium-Range Weather Forecasts, Technical Memorandum 787.

Hogan, R. J., & Illingworth, A. J. (2000). Deriving cloud overlap statistics from radar. *Quarterly Journal of the Royal Meteorological Society*, *126*(569), 2903–2909.

Hogan, R. J., Schäfer, S. A., Klinger, C., Chiu, J. C., & Mayer, B. (2016). Representing 3-D cloud radiation effects in two-stream schemes: 2. Matrix formulation and broadband evaluation. *Journal of Geophysical Research: Atmospheres*, *121*, 8583–8599. https://doi.org/10.1002/2016JD024876

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research*, *113*, D13103. https://doi.org/10.1029/2008JD009944

Kalmus, P., Lebsock, M., & Teixeira, J. (2014). Observational boundary layer energy and water budgets of the stratocumulus-to-cumulus transition. *Journal of Climate*, *27*(24), 9155–9170.

Lebsock, M. D., L'Ecuyer, T. S., & Stephens, G. L. (2011). Detecting the ratio of rain and cloud water in low-latitude shallow marine clouds. *Journal of Applied Meteorology and Climatology*, *50*(2), 419–432.

Lebsock, M., Morrison, H., & Gettelman, A. (2013). Microphysical implications of cloud-precipitation covariance derived from satellite remote sensing. *Journal of Geophysical Research: Atmospheres*, *118*, 6521–6533. https://doi.org/10.1002/jgrd.50347

Lebsock, M., & Su, H. (2014). Application of active spaceborne remote sensing for understanding biases between passive cloud water path retrievals. *Journal of Geophysical Research: Atmospheres*, *119*, 8962–8979. https://doi.org/10.1002/2014JD021568

Lewis, E. R., Wiscombe, W. J., Albrecht, B. A., Bland, G. L., Flagg, C. N., Klein, S. A., et al. (2012). MAGIC: Marine ARM GPCI investigation of clouds (*Tech. rep. DOE/SC-ARM-12-020*). United States: DOE Office of Science Atmospheric Radiation Measurement (ARM) Program (United States).

Li, J.-L., Waliser, D., Stephens, G., Lee, S., L'Ecuyer, T., Kato, S., et al. (2013). Characterizing and understanding radiation budget biases in CMIP3/CMIP5 GCMs, contemporary GCM, and reanalysis. *Journal of Geophysical Research: Atmospheres*, *118*, 8166–8184. https://doi.org/10.1002/jgrd.50378

Li, J.-L. F., Waliser, D., Woods, C., Teixeira, J., Bacmeister, J., Chern, J., et al. (2008). Comparisons of satellites liquid water estimates to ECMWF and GMAO analyses, 20th century IPCC AR4 climate simulations, and GCM simulations. *Geophysical Research Letters*, *35*, L19710. https://doi.org/10.1029/2008GL035427

Lin, J.-L., Qian, T., & Shinoda, T. (2014). Stratocumulus clouds in Southeastern Pacific simulated by eight CMIP5–CFMIP global climate models. *Journal of Climate*, *27*(8), 3000–3022.

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., et al. (2018). Clouds and the Earth's radiant energy system (CERES) energy balanced and filled (EBAF) top-of-atmosphere (TOA) edition-4.0 data product. *Journal of Climate*, *31*(2), 895–918.

Mace, G. G., & Benson-Troth, S. (2002). Cloud-layer overlap characteristics derived from long-term cloud radar data. *Journal of Climate*, *15*(17), 2505–2515.

Martin, G., Johnson, D., & Spice, A. (1994). The measurement and parameterization of effective radius of droplets in warm stratocumulus clouds. *Journal of the Atmospheric Sciences*, *51*(13), 1823–1842.

McGibbon, J., & Bretherton, C. (2017). Skill of ship-following large-eddy simulations in reproducing MAGIC observations across the northeast Pacific stratocumulus to cumulus transition region. *Journal of Advances in Modeling Earth Systems*, *9*, 810–831. https://doi.org/10.1002/2017MS000924

Minnis, P., Sun-Mack, S., Chen, Y., Khaiyer, M. M., Yi, Y., Ayers, J. K., et al. (2011). CERES edition-2 cloud property retrievals using TRMM VIRS and Terra and Aqua MODIS data—Part II: Examples of average results and comparisons with other data. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(11), 4401–4430.

Minnis, P., Sun-Mack, S., Young, D. F., Heck, P. W., Garber, D. P., Chen, Y., et al. (2011). CERES edition-2 cloud property retrievals using TRMM VIRS and Terra and Aqua MODIS data—Part I: Algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(11), 4374–4400.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research*, *102*(D14), 16,663–16,682. https://doi.org/10.1029/97JD00237

Morcrette, J., Barker, H., Cole, J., Iacono, M., & Pincus, R. (2008). Impact of a new radiation package, McRad, in the ECMWF integrated forecasting system. *Monthly Weather Review*, *136*(12), 4773–4798.

Nam, C., Bony, S., Dufresne, J.-L., & Chepfer, H. (2012). The too few, too bright tropical low-cloud problem in CMIP5 models. *Geophysical Research Letters*, *39*, L21801. https://doi.org/10.1029/2012GL053421

Neggers, R. A., Heus, T., & Siebesma, A. P. (2011). Overlap statistics of cumuliform boundary-layer cloud fields in large-eddy simulations. *Journal of Geophysical Research*, *116*, D21202. https://doi.org/10.1029/2011JD015650

Oreopoulos, L., & Khairoutdinov, M. (2003). Overlap properties of clouds generated by a cloud-resolving model. *Journal of Geophysical Research*, *108*(D15), 4479. https://doi.org/10.1029/2002JD003329

Painemal, D., Greenwald, T., Cadeddu, M., & Minnis, P. (2016). First extended validation of satellite microwave liquid water path with ship-based observations of marine low clouds. *Geophysical Research Letters*, *43*, 6563–6570. https://doi.org/10.1002/2016GL069061

Painemal, D., & Zuidema, P. (2011). Assessment of MODIS cloud effective radius and optical thickness retrievals over the Southeast Pacific with VOCALS-REx in situ measurements. *Journal of Geophysical Research*, *116*, D24206. https://doi.org/10.1029/2011JD016155

Platnick, S. (2000). Vertical photon transport in cloud remote sensing problems. *Journal of Geophysical Research*, *105*(D18), 22,919–22,935. https://doi.org/10.1029/2000JD900333

Platnick, S., King, M. D., Ackerman, S. A., Menzel, W. P., Baum, B. A., Riédi, J. C., & Frey, R. A. (2003). The MODIS cloud products: Algorithms and examples from Terra. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(2), 459–473.

Schäfer, S. A., Hogan, R. J., Klinger, C., Chiu, J. C., & Mayer, B. (2016). Representing 3-D cloud radiation effects in two-stream schemes: 1. Longwave considerations and effective cloud edge length. *Journal of Geophysical Research: Atmospheres*, *121*, 8567–8582. https://doi.org/10.1002/2016JD024875

Seethala, C., & Horváth, Á. (2010). Global assessment of AMSR-E and MODIS cloud liquid water path retrievals in warm oceanic clouds. *Journal of Geophysical Research*, *115*, D13202. https://doi.org/10.1029/2009JD012662

Shonk, J. K., & Hogan, R. J. (2008). Tripleclouds: An efficient method for representing horizontal cloud inhomogeneity in 1D radiation schemes by using three regions at each height. *Journal of Climate*, *21*(11), 2352–2370.

Shonk, J. K., Hogan, R. J., Edwards, J. M., & Mace, G. G. (2010). Effect of improving representation of horizontal and vertical cloud structure on the Earth's global radiation budget. Part I: Review and parametrization. *Quarterly Journal of the Royal Meteorological Society*, *136*(650), 1191–1204.

Teixeira, J., Cardoso, S., Bonazzola, M., Cole, J., DelGenio, A., DeMott, C., et al. (2011). Tropical and subtropical cloud transitions in weather and climate prediction models: The GCSS/WGNE Pacific cross-section intercomparison (GPCI). *Journal of Climate*, *24*(20), 5223–5256.

Tompkins, A. M., & Di Giuseppe, F. (2015). An interpretation of cloud overlap statistics. *Journal of the Atmospheric Sciences*, *72*(8), 2877–2889.

Vial, J., Dufresne, J.-L., & Bony, S. (2013). On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates. *Climate Dynamics*, *41*(11-12), 3339–3362.

Webb, M. J., Lambert, F. H., & Gregory, J. M. (2013). Origins of differences in climate sensitivity, forcing and feedback in climate models. *Climate Dynamics*, *40*(3-4), 677–707.

Wentz, F. J. (2013). SSM/I version-7 calibration report (Techical Report 11012, 46.) Santa Rosa, CA: Remote Sensing Systems.

Wentz, F. J., & Spencer, R. W. (1998). SSM/I rain retrievals within a unified all-weather ocean algorithm. *Journal of the Atmospheric Sciences*, *55*(9), 1613–1627.

Winker, D. M., Pelon, J. R., & McCormick, M. P. (2003). The CALIPSO mission: Spaceborne lidar for observation of aerosols and clouds. In *Third International Asia-Pacific Environmental Remote Sensing Remote Sensing of the Atmosphere, Ocean, Environment, and Space, International Society for Optics and Photonics* (pp. 1–11).

Wood, R. (2000). Parametrization of the effect of drizzle upon the droplet effective radius in stratocumulus clouds. *Quarterly Journal of the Royal Meteorological Society*, *126*(570), 3309–3324.

Zhang, Z., Dong, X., Xi, B., Song, H., Ma, P.-L., Ghan, S. J., et al. (2017). Intercomparisons of marine boundary layer cloud properties from the ARM CAP-MBL campaign and two MODIS cloud products. *Journal of Geophysical Research: Atmospheres*, *122*, 2351–2365. https://doi.org/10.1002/2016JD025763

Zhou, X., Kollias, P., & Lewis, E. R. (2015). Clouds, precipitation, and marine boundary layer structure during the MAGIC field campaign. *Journal of Climate*, *28*(6), 2420–2442.

Zuidema, P., & Joyce, R. (2008). Water vapor, cloud liquid water paths, and rain rates over northern high latitude open seas. *Journal of Geophysical Research*, *113*, D05205. https://doi.org/10.1029/2007JD009040